Open Journal of Genetics, 2012, 2, 5-10 **OJGen**
Published Online December 2012 (http://www.SciRP.org/journal/ojgen/)

Organization, expression and evolution of flagellar genesin *Rhodobacter sphaeroides* 2.4.1

Durga Thapaliya[1],B. Myagmarjav[1], C. Trahan[1],D. Ortiz[1],H. Cho[2], M. Choudhary[1*]

[1]Department of Biological Sciences, Sam Houston State University, Huntsville, Texas, USA
[2]Department of Computer Science, Sam Houston State University, Huntsville, Texas, USA
Email: *mchoudhary@shsu.edu

Received 2012

ABSTRACT

***Rhodobacter sphaeroides* 2.4.1 belongs to theα-3 subdivision of the *Proteobacteria*. It possesses a multipartite genome structure consisting of two circular chromosomes, andit displays a wide range of metabolic diversity.Approximately 40 flagellar proteins are required for structure, assembly, and regulation of the flagellum formation in most bacterial species. *R. sphaeroides*contains two flagellar gene clusters (fla1 and fla2),which encode 38 and 21 proteins, respectively. Thirty-six of these genes exist in duplicate gene-pairs.A combination of genome analysis, phylogenetic analysis and mRNA expression analysis were employed to examine the conservation of structure, function and evolution of fla1 and fla2 in *R. sphaeroides*. The results demonstrated that fla2, which was shared among members of α-*Proteobacteria*, is native to*R. sphaeroides*, while fla1 was horizontally transferred from a member of γ-*Proteobacteria*.In addition, genes located in fla1 are expressed over several growth conditions, but those in fla2 are barely expressed.**

Keywords: Flagella; Horizontal Gene Transfer; Phylogenetic Tree

1. INTRODUCTION

Bacterial flagella are complex structures that facilitate different types of motilities (swimming, swarming, gliding and twitching), and play important roles in sensing outside environments (temperature, nutrient and oxygen availability), adhesion, biofilm production, and host invasion [1]. A bacterial flagellum is composed of at least 21-24 core proteins [2, 3], which represent six structural components of the flagellum, including a basal body (MS ring, P ring, and L ring), a motor, a switch, a hook, a filament, and an export apparatus.In addition, another set of 15-25 proteins is responsible for the regulation of flagellar assembly and the uncovering and processing of environmental signals to which flagella respond [4].

A large number of bacterial species contain a single flagellar gene cluster, which providescells with different types of motilities. However, a number of bacterial species inthe genera*Vibrio*,*Rhodospirillum*, *Bradyrhizobium, Burkholderia*, and *Yersinia*possess twoflagellar gene clusters [5], which contain genes that encode proteins for the synthesis ofpolar and lateral flagella to control swimming and swarming, respectively. Although the ability to synthesize different flagella types is primarily encoded by structural genes, the variation in regulation mechanisms providesthese microorganisms varied strategies to exploit a diverse range of ecological niches.

R. sphaeroides is a purple non-sulfur photosynthetic bacterium, which belongs to the α-3 subgroup of the*Proteobacteria* [6].The genome of *R. sphaeroides* consists of two circular chromosomes [7], which has been completely sequenced and annotated [8]. It also exhibits a prevalence of gene duplications [9, 10]. Genome analysis revealed that the primary chromosome of R. sphaeroides contains two flagellar gene clusters, fla1 (between 1,736,242 and 1,951,757 base-pairs) and fla2 (between 3,074,540 and 3,105,787 base-pairs). It has been found that the fla1 cluster contains 38genes, and is responsible for the formation ofthe polar flagellum, while fla2 cluster contains 21 genes, whose functions remain unclear.

The duplicate gene-pairs that exist between fla1 and fla2 clusters in *R. sphaeroides*may have resulted from either segmental gene duplicationor horizontal gene transfer [11]. The two gene clusters may have diverged since gene duplication or horizontal gene transfer, andthe two gene clusters would have evolved and expressed differently under different growth conditions. To study structure and function of these duplicate flagellargenes, current study employsthe following approaches, includingsequence analysis, phylogenetic

*Corresponding author.

analysis, and mRNA expression analysis.

Table1. Similarity of duplicate genes between fla1 and fla2 clusters in *R. sphaeroides* 2.4.1.

Gene[a]	RSP No.[b]	Coverage[c]	Score[d]	Identity[e]	E-value[f]
flhA1 / flhA2	0034 / 1320	97	364	36	3.00E-117
flgI1 / flgI2	0076 / 1307	93	249	43	8.00E-81
flgG1 / flgG2	0078 / 1326	96	207	42	1.00E-67
fliP1 / fliP2	0063 / 1309	90	176	40	2.00E-55
flhB1 / flhB2	0066 / 1322	94	136	33	2.00E-38
flgE1 / flgE2	0080 / 1303	98	108	25	4.00E-28
fliF1 / fliF2	0053 / 1312	77	103	30	1.00E-25
fliR1 / fliR2	0065 / 1321	85	77	33	8.00E-19
flgF1 / flgF2	0079 / 1327	90	77	32	4.00E-19
flgH1 / flgH2	0077 / 1324	74	75.1	30	7.00E-19
fliQ1 / fliQ2	0064 / 1328	79	62.8	46	3.00E-16
flgD1 / flgD2	0081 / 1336	30	62.4	46	3.00E-14
flgK1 / flgK2	0074 / 1304	64	46.6	37	3.00E-07
motA1 / motA2	0233 / 1316	87	43.5	22	1.00E-07
flgA1 / flgA2	0036 / 1325	70	43.1	33	8.00E-08
flgC1 / flgC2	0082 / 1330	83	41.2	43	7.00E-08
flgB1 / flgB2	0083 / 1331	71	40	32	2.00E-07
fliE1 / fliE2	0052 / 1329	100	32	28	1.00E-04

[a]Gene name reflecting cluster 1 (fla1) and 2 (fla2),[b]*R. sphaeroides* gene number,[c]Percentage query coverage,[d]Normalized score,[e]Percentage amino acid identity,[f]Expected value

2. MATERIALS AND METHODS

2.1. Sequence Analysis

Identification of a gene homolog within the genome of *R. sphaeroides*was performed using gapped BLASTP [12], where genes from fla1 were used as queries to identify the corresponding homolog in the fla2 cluster. Each copy of the duplicate gene-pair was also used to identifyorthologs among bacterial species representing different groups of *Proteobacteria*, using the symmetrical best-hit method. To identify the orthologs, three criteria (E-score threshold $<10^{-3}$, query coverage of 50%, and overall amino acid identity >30%) were used.All duplicate genes were analyzed for their %GC content, di- and tri-nucleotide repeat patterns, and codon usage.

2.2. Global DNA Sequence Alignment

The flagellar gene clusters were identified in the genomic sequences of *R. sphaeroides* 2.4.1, *Ruegeriapomeroyi* DSS-3, and *Pseudoxanthomonassuwonensis* 11-1. DNA sequence files (in fasta format)of each flagellar cluster were downloaded from the NCBI database and the sequence alignments were performed using Mauve 2.3.1 [13]. Number of locally collinear blocks (LCBs) and % conservation of the DNA regionsweredetermined as previously described [9].

2.3. Phylogenetic Analysis

Phylogenetic analysis was performed for the five coregenes (flhA, flgG, fliP,flhB, and flgE),which were selected with the criteria of percentage query coverage >90% and normalized score >100. Geneious v4.6 was used to generate alignments using MUSCLE [14] and construct phylogenetic treesusing Neighbor-Joining (NJ) [15] and Maximum-Likelihood (ML)methods [16]with 100 bootstrap replications. Protein sequences of these corresponding genes were obtained from a total of 75 proteobacterial species (15α-, 15 β-, 28 γ-, 8 δ-, and 9 ε-) for phylogenetic tree construction.

2.4. Expression Analysis

Microarray expression data of*R. sphaeroides*2.4.1 were available [17], from which the expression data of the flagellar genes were collected and used for this study. The expression levels were measured under seven growth conditions, includingthree photosynthetic conditions (3 watts, 10 watts, and 100 watts), aerobic, semi-aerobic, and dark-dimethyl-sulfoxide(DMSO). Individual and average expression levels of genes in fla1 and fla2 were compared.

3. RESULTS

3.1. Identification of Flagellar Gene Homologs in *R. sphaeroides* genome

A majority offlagellar genes are located in the two gene clusters, fla1 (between 1,736,242 and 1,951,757 base-pairs) and fla2 (between 3,074,540 and 3,105,787 base-pairs) on the primary chromosome of R. sphaeroides as previously designated [11].The fla1 and fla2 gene clusters encode 38 and 21 proteins, respectively. A singleton gene (*fliG2*) is located in a different region (between 831,570 and 832,643 base-pairs)of the same chromosome. In addition, a mini-cluster containing four genes (*flaA*, *flaF*, *flbT*, and *flgF*) are located in plasmid pRS241A. The resultsof the amino acid similarity between 18 gene-pairs wereshown in **Table 1**.The amino acid identities for these gene-pairs range from 22% to

Table 2. Comparison of genomic alignments of the flagellar gene clusters.

Aligned flagellar clusters	No.LCBs[a]	LCBLength[b]	% Conserved[c]
R.sphaeroides-fla1(α)/*R.sphaeroides-fla2*(α)	23	41065	13.02
R. sphaeroides-fla1(α) / *R. pomeroyi*(α)	18	34693	8.38
R. sphaeroides-fla1(α) / *P. suwonensis*(γ)	21	68487	16.08
R. sphaeroides-fla2(α) / *R. pomeroyi*(α)	5	44714	25.78
R. sphaeroides-fla2(α) / *P. suwonensis*(γ)	11	41542	8.01
R.pomeroyi(α) / *P. suwonensis*(γ)	19	31978	16.07

[a]localcolinear block (LCB), [b]LCB length (in base-pair), [c]LCB percentage (over the total length of gene cluster)

46%. Six of the 18 gene-pairs exhibit similarity >90% of their corresponding protein lengths with >100 normalized scores. Furthermore, the genes located in the fla1cluster also indicated significant matches to the members of the γ-*Proteobacteria*, and the amino acid identity between the two copies of each corresponding ortholog were higher than the amino acid identity shown between the corresponding duplicate copies withinR. sphaeroides. In contrast, genes in the fla2 cluster mostly match to the corresponding homologs ofα-proteobacterial species (data not shown).

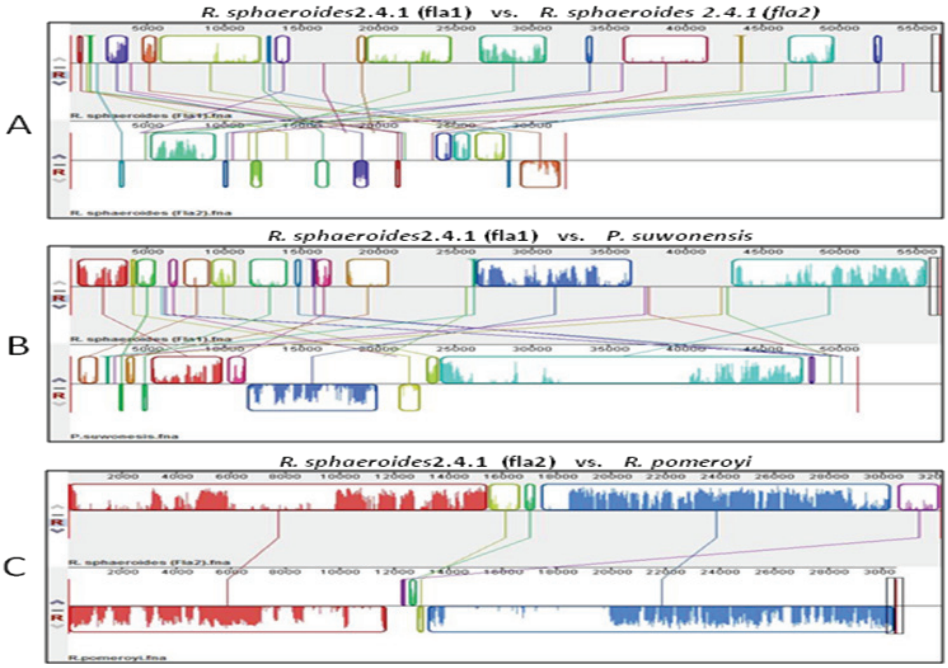

Figure 1. Mauve representation of flagellar clusters of *R. sphaeroides* 2.4.1, *P. suwonensis*, and *R. pomeroyi*. Each sequential color-block represents a homologous backbone DNA sequence without rearrangement.

3.2. Comparison of fla1 and fla2

Comparisons of all pairwise DNA sequence alignments of the fla1 and fla2 regions of *R. sphaeroides*, *R. pomeroyi,* and *P. suwonensis* were described in **Table 2**. Thealignments between fla1and fla2 of R. sphaeroides, between fla1 of R. sphaeroides and the flagellar region of P. suwonensis, and between fla2 of *R. sphaeroides* and the flagellar region of*R. pomeroyi* were shown in **Figure 1A**, **1B**, and **1C**, respectively. As shown in **Table 2**, fla2 of R. sphaeroides and the flagellar gene cluster of *R.pomeroyi*shared five LCBs with 25.78% sequence conservation.The organization of LCBs in thegenomic regionis indicative of two large scale inversions in R. pomeroyi as shown in **Figure 1C**. However, fla1 and fla2 of *R. sphaeroides* shared 23 LCBs with the low level (13.02%)of their sequence conservation, which is also demonstrated by small LCBs with a large number of chromosomal rearrangements as shown in **Figure 1A**. The fla1 cluster of *R. sphaeroides*and the flagellar cluster of *P.suwonensis*11-1 shared 21LCBs with 16.07% sequence conservation, and this medium level conservation is also reflected in larger LCBs as shown in **Figure 1B**.

The five common genes in fla1 and fla2 clusters of *R. sphaeroides*, and their corresponding homologs in *R. pomeroyi* and *P. suwonensis* were analyzed for the %GC content, di- and tri-nucleotide repeats, and codon frequencies. The % GC content, di- and tri-nucleotide repeat, and codon frequency distributions are similar among these gene homologs (data not shown).

3.3. Phylogenetic Analysis

Phylogenetic analysis based on each of the fiveprotein (FlhA, FlgG, FliP, FlhB, and FlgE)trees reflected a similar evolutionary relationship among 75 proteobacterial species. Two FlgG trees based on NJ distance and maximum-likelihood methodswere shown in **Figure 2A** and **2B**, respectively. The results revealed that genes located in fla2 of *R. sphaeroides* form a clade with its closely related species, *R. pomeroyi* and *P. denitrificans* with a bootstrap value of 100. All three species belong to the order *Rhodobacteriales*, and are located within the branch of α-*Proteobacteria*. In contrast, the genes lo-

